国金证券 SINOLINK SECURITIES

2025 年 03 月 23 日

AI 周观察

数据专题研究

证券研究报告

分析师：刘道明（执业 S1130520020004） liudaoming@gjzq.com.cn
联系人：黄晓军（执业 S1130122050092） huangxiaojun@gjzq.com.cn
联系人：麦世学（执业 S1130123100111） maishixue@gjzq.com.cn

英伟达下一代产品面积进一步增大，Gemini 持续发布新功能

摘要

- 英伟达在 GTC2025 发布了 Blackwell Ultra、Rubin 和 Rubin Ultra 加速卡，显示未来两代加速卡面积进一步增大。Rubin 面积约为两倍光罩极限，FP4 算力为 Blackwell Ultra 三倍；Rubin Ultra 面积达四倍光罩极限，算力进一步翻倍。由于单颗中介层面积逼近八倍光罩极限，Rubin Ultra 采用两块中介层+I/O die 的封装设计，以超大型 ABF 基板替代传统大尺寸中介层，折射出当前大面积 CoWoS 封装的技术挑战。
- 美光科技 FY25 Q2 财报显示，公司本季度营收 80.53 亿美元，同比增长 38.27%，环比下滑 7.53%。其中，DRAM 收入环比降 4%，受 bit 出货量下降抵消价格上涨影响；NAND 收入环比降 17%，因低端消费级产品占比提升。公司存储业务（SBU）收入环比下滑 20%，反映数据中心客户采购回落。尽管 NAND 价格近期上涨，公司维持较低稼动率，市场景气度能否持续回暖仍待观察。
- 英伟达在 GTC2025 上提出 Agentic AI 概念，旨在通过多模型协同工作提升 AI 能力，叠加思考时的计算量，预计将显著提升对加速卡存储容量与带宽的需求。Rubin Ultra 单卡存储容量达 1024GB，带宽提升至 32TB/s。与此同时，SK hynix 发布的 12 层 HBM4 突破 2TB/s 带宽大关，较 HBM3E 提升 60%，并采用 MR-MUF 工艺提升散热与稳定性。我们认为，随着 Agentic AI 的兴起，HBM 领域将持续维持较高景气度。
- 聊天助手类应用海外市场整体稳定，Perplexity 受竞品冲击访问量下降；国内市场多数应用微降，文心一言因新模型发布访问量大幅增长。Google Gemini 推出“画布”与音频概览功能增强用户体验。OpenAI 发布三款语音模型，提升语音识别与合成能力。Mistral AI 和腾讯分别发布高性能开源模型。Nvidia 的 Dynamo 软件大幅提升 DeepSeek 的 AI 处理速度。HPC-AI Tech 的 Open-Sora 2.0 以低成本实现高性能，Stability AI 推出 2D 转沉浸式视频工具。腾讯混元和 Roblox 分别发布新的 3D 生成模型，提升 3D 内容创作效率并降低成本。
- 当前自动驾驶痛点包括训练数据不足、决策规划的安全性不足、传感器精度不足、系统安全性不足、车载芯片算力不足、高精地图缺失等等。英伟达 Omniverse+Cosmos 方案下可以帮助解决训练数据、决策规划安全性的不足，Halos 将保护 Hyperion 系统级的安全，新一代 Thor 芯片也将解决短期内芯片算力不足的情况。我们认为自动驾驶发展将会进一步加速，与英伟达有合作的企业将会率先受益。除了合成数据帮助训练自动驾驶算法之外，真实世界驾驶数据也将帮助算法训练。国内安防企业通过摄像头已采集大量车辆驾驶数据，我们认为这些公司也将受益。
- 目前在机器人制造方面，国内产业链更为成熟，企业产能、供应链管理相比美国企业更具优势，但在机器人大脑、小脑智能上相比有英伟达芯片支持的美国企业仍有差距。我们认为英伟达 Omniverse+Cosmos+Groot N1 生态叠加 Thor 平台今年上线将会加速机器人具身智能发展，看好产业链受益。

风险提示

芯片制程发展与良率不及预期

中美科技领域政策恶化

智能手机销量不及预期

内容目录

海外市场行情回顾

图表1：截至 3 月 21 日海外 AI 相关个股行情

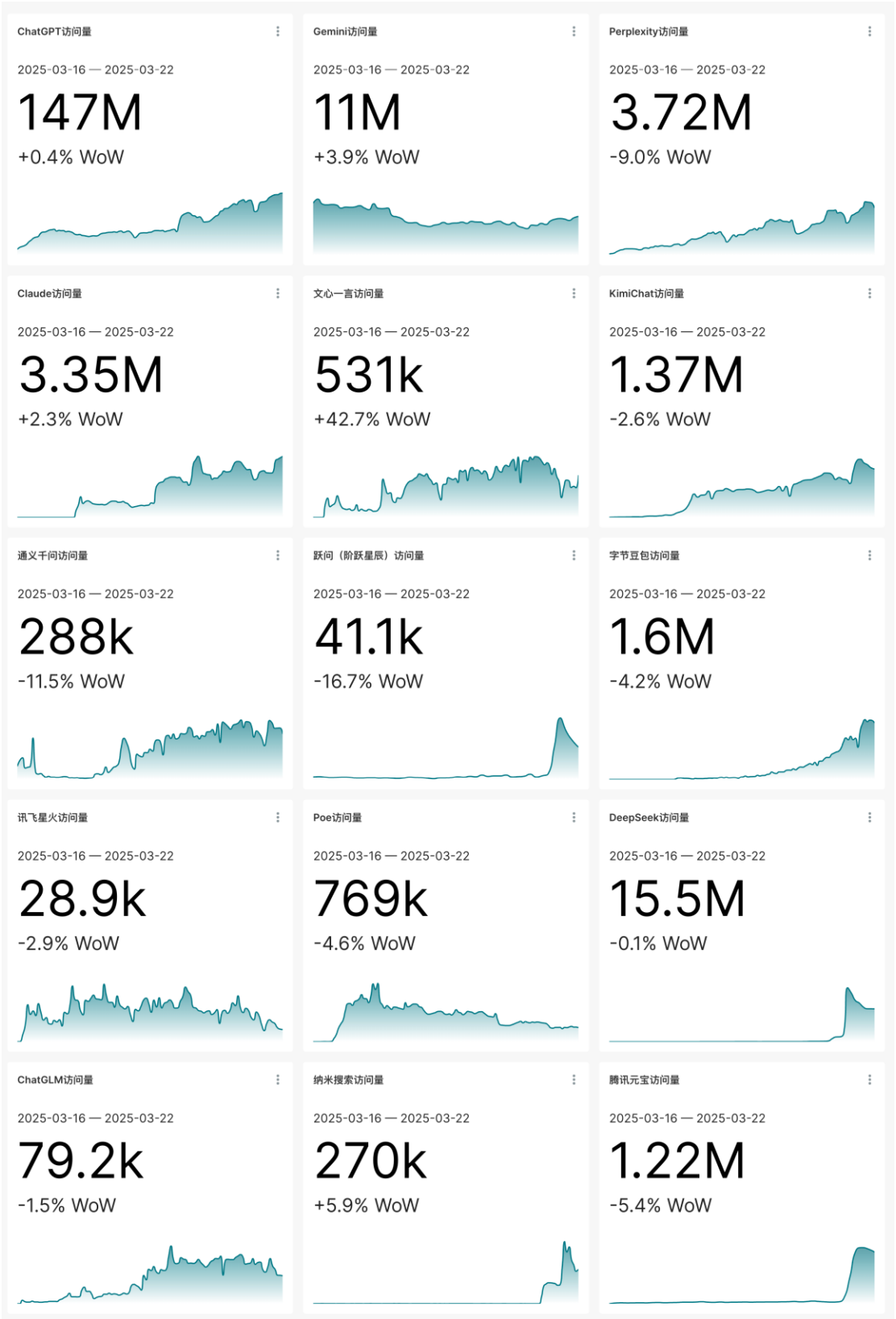
个股名称	个股代码	本周收盘价	上周收盘价	涨跌幅	类目
Palantir	PLTR	90.96	86.24	5.47%	AI模型与应用
超威半导体	AMD	106.44	100.97	5.42%	AI算力
C3.AI	AI	22.62	21.61	4.67%	AI模型与应用
Mongodb	MDB	192.54	185.37	3.87%	AI模型与应用
Zscaler	ZS	205.2	197.81	3.74%	AI模型与应用
Zeta	ZETA	14.43	13.98	3.22%	AI模型与应用
DataDog	DDOG	105.03	101.8	3.17%	AI模型与应用
Cloudflare	NET	119.22	116.15	2.64%	AI模型与应用
CrowdStrike	CRWD	362.24	353.735	2.40%	AI模型与应用
美满	MRVL	70.39	68.74	2.40%	AI算力
苹果	AAPL.O	218.27	213.49	2.24%	消费电子和汽车
戴尔	DELL.N	97.57	95.67	1.99%	消费电子和汽车
Snowflake	SNOW	158.39	156.11	1.46%	AI模型与应用
惠普	HPQ.N	28.68	28.41	0.95%	消费电子和汽车
英特尔	INTC.O	24.26	24.05	0.87%	消费电子和汽车
微软	MSFT	391.26	388.56	0.69%	AI模型与应用
Qorvo	QRVO.O	71.8	71.39	0.57%	消费电子和汽车
Salesforce	CRM	280.62	279.4	0.44%	AI模型与应用
高通	QCOM.O	156.82	156.58	0.15%	消费电子和汽车
Palo Alto	PANW	182.32	182.34	-0.01%	AI模型与应用
特斯拉	TSLA.O	248.71	249.98	-0.51%	消费电子和汽车
亚马逊	AMZN	196.21	197.95	-0.88%	AI模型与应用
谷歌	GOOGL	163.99	165.49	-0.91%	AI模型与应用
Meta	META	596.25	607.6	-1.87%	AI模型与应用
博通	AVGO	191.66	195.54	-1.98%	AI算力
Gitlab	GTLB	50.95	52.08	-2.17%	AI模型与应用
Skyworks	SKWS.O	67.25	69.4	-3.10%	消费电子和汽车
英伟达	NVDA	117.7	121.67	-3.26%	AI算力
Innodata	INOD	41.82	48.18	-13.20%	AI模型与应用

来源：Reuters、国金证券研究所

AI 模型与应用动态

“小”模型竞争激烈，Google AI 应用持续更新新功能

图表2：聊天助手类 AI 应用活跃度



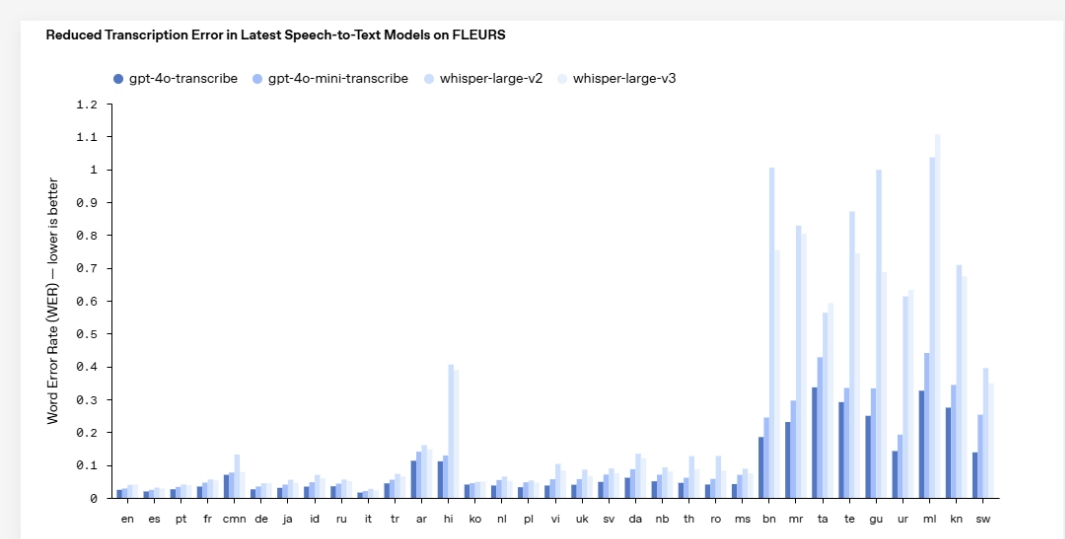
来源：Similarweb、国金证券研究所

从聊天助手类应用访问量看，海外应用整体环比变化不大，Perplexity 环比下降 9%，可能是受到各大应用厂商推出 Deep Research 的冲击。国内应用多数环比微降，文心一言受益于新模型的发布，访问量环比上升超过 40%。

Google Gemini 推出“画布”与音频概览功能，提升用户协作与内容消费体验。“画布”功能提供互动空间，方便用户进行写作与编程项目的创作、改进与分享；音频概览功能则可将文档、网页等内容生成播客形式的音频摘要，目前仅支持英文。

OpenAI 发布三款全新语音模型：gpt-4o-transcribe、gpt-4o-mini-transcribe 及 gpt-4o-mini-tts。gpt-4o-transcribe 作为 Whisper 升级版，在 33 种语言测试中错误率显著降低，定价与 ElevenLabs Scribe 持平（每百万音频输入 tokens 6 美元）。gpt-4o-mini-tts 为文本转语音模型，支持多种音色、语速及风格。

图表3：gpt-4o-transcribe 不同语言评分

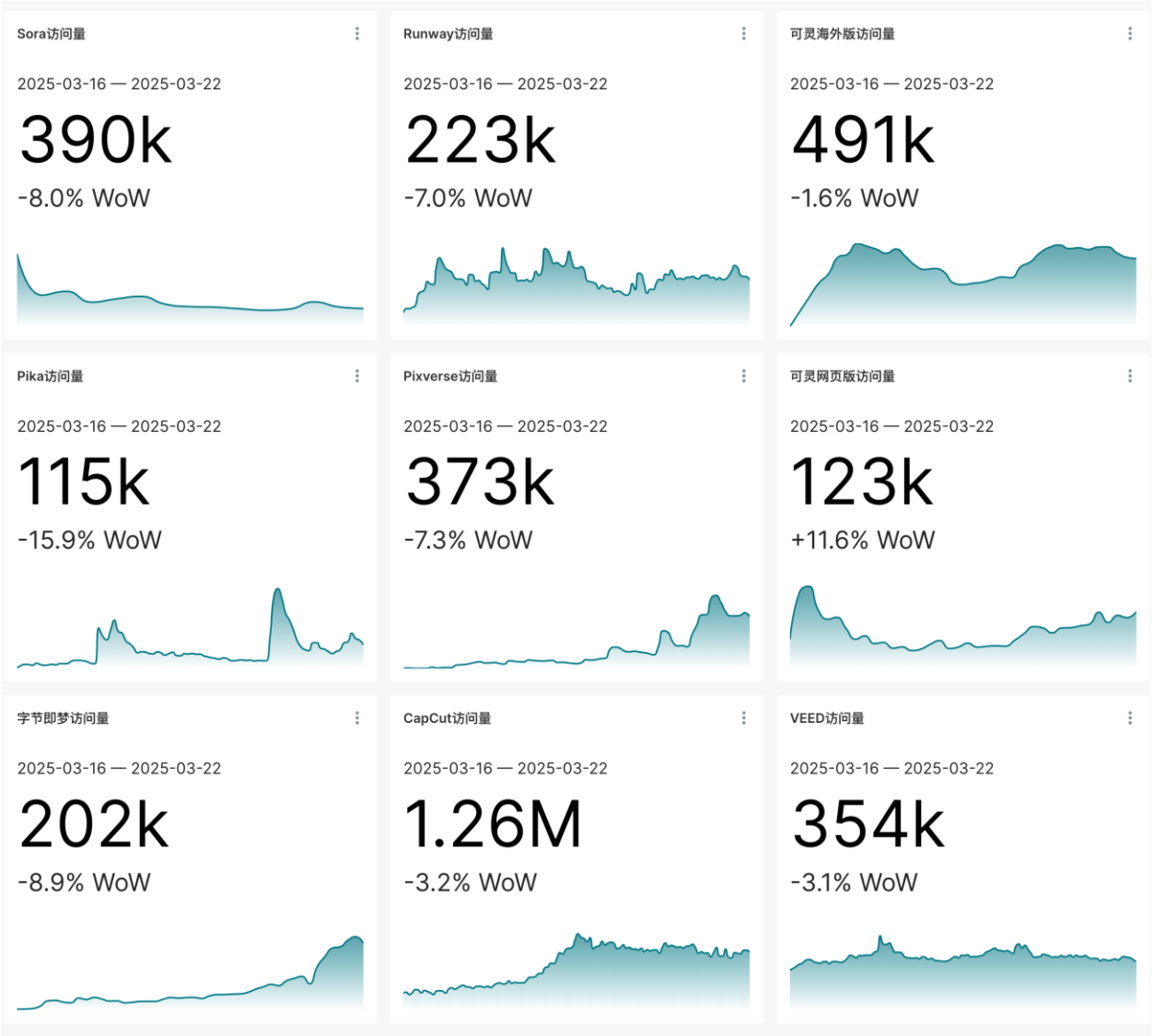


来源：OpenAI、国金证券研究所

Mistral AI 发布开源模型 Mistral Small 3.1（24B 参数），性能超越 Gemma 3 27B 及 GPT-4o mini，支持多模态理解和 128k tokens 上下文窗口。腾讯推出混元 T1 正式版，沿用 Turbo S 的 Hybrid-Mamba-Transformer 架构，降低了计算复杂度与 KV-Cache 内存占用，优化训练与推理成本。

在 3 月 18 日的 GTC 大会上，Nvidia 首席执行官黄仁勋宣布推出 Dynamo 软件，旨在将 DeepSeek 的 AI 处理速度提升 30 倍。Dynamo 软件能够将 AI 推理任务分配到多达 1000 个 GPU 上并行处理，显著提升查询吞吐量，服务提供商能够更高效地处理客户查询，从而提高收入。

图表4：视频生成类AI应用活跃度



来源：Similarweb、国金证券研究所

HPC-AI Tech 推出 Open-Sora 2.0，训练成本仅 20 万美元，远低于同类系统。采用三阶段训练和视频 DC-AE 自动编码器，实现 5.2 倍训练加速和超 10 倍生成加速。VBench 得分与 Sora 仅差 0.69%，视觉质量与提示准确性表现出色。Stability AI 发布 Stable Virtual Camera，可将 2D 图像转换为具真实深度和视角的沉浸式视频。

腾讯混元发布五个全新开源 3D 生成模型，基于 Hunyuan3D-2.0，生成速度更快、细节更丰富。Turbo 系列模型利用 FlashVDM 框架实现 30 秒内生成。升级后的 3D AI 创作引擎支持多视图输入，通过少量图片即可生成高质量 3D 模型，降低制作成本。新模型适用于 UGC、商品素材合成及游戏资产生成。Roblox 开源其首个 3D 对象生成基础模型 Cube3D，通过创新训练方法将 3D 对象标记化，实现快速生成完整 3D 形状，提升 3D 创作效率。

Rubin 和 Rubin Ultra 芯片面积进一步增大，封装难度进一步提升

英伟达于 GTC2025 更新了其加速卡产品线，发布了本世代 Blackwell 架构 HBM 加强版 Blackwell Ultra 以及次世代架构的 Rubin 和 Rubin Ultra 加速卡，值得注意的是，从官方披露的加速卡图片来看，未来两代加速卡面积进一步增大。

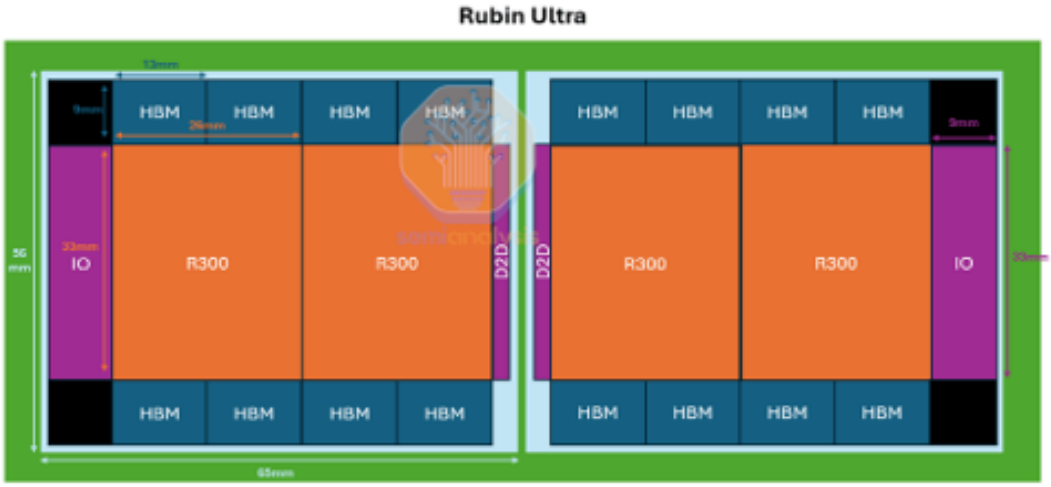
图表5：Rubin 和 Rubin Ultra 芯片面积进一步增大

来源：英伟达、国金证券研究所

从示意图上来看，尽管官方标注 Rubin 面积约为两倍光罩极限，但略宽于 Blackwel Ultra，算力性能层面，Rubin FP4 算力指标约为 Blackwell Ultra 三倍，推测 Rubin 使用了比 N4P 更为先进的制程 N3P，而 Rubin Ultra 面积官方标注为四倍光罩极限，相应的算力性能翻倍，可推测 Rubin Ultra 采用了和 Rubin 相同的制程，性能的增长来自于 die size 的扩大。

更大的面积带来的将是封装难度的进一步提升，Semianalysis 预计 Rubin Ultra 将采用两块中介层组成的封装结构，以避免使用一块超大型中介层，因为当 GPU die size 达到四倍光罩极限后，叠加 16 块 HBM 堆栈后，中介层的面积逼近八倍光罩极限，这种超大尺寸已接近当前封装技术的极限。在 Rubin Ultra 的封装设计中，两颗位于中间的 GPU die 之间将通过一颗 I/O die 进行同通信。GPU die 之间的通信将通过更下层的基板(Substrate)而非传统的大型中介层来完成。这种设计将采用一块超大的 ABF 基板（Ajinomoto Build-up Film），其尺寸将超出当前 JEDEC 标准封装的限制（最大为 120mm × 120mm），对制造、封装以及散热设计提出了更高的挑战

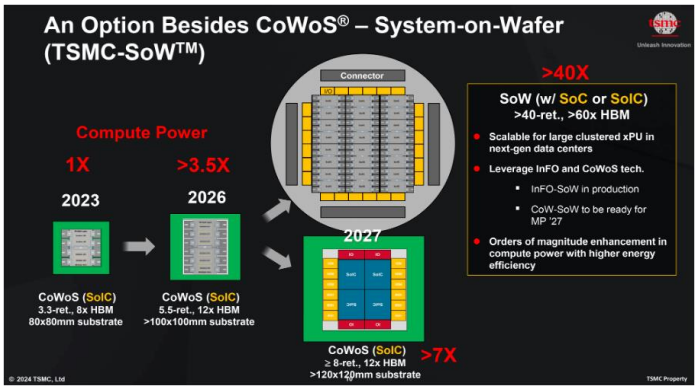
图表6：Rubin Ultra 封装示意图



来源：Semianalysis、国金证券研究所

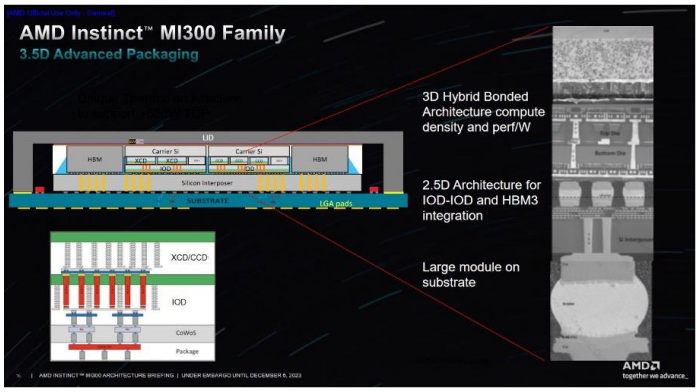
此前我们曾提出，在后摩尔时代，先进封装将成为弥补晶体管缩放速率下滑的重要手段。然而，这一技术路线同样面临瓶颈。其中，CoWoS 中介层面积的受限便是制约单芯片性能进一步提升的关键因素之一。根据台积电于 2024 年 3 月发布的路线图，CoWoS 中介层面积预计将在 2027 年达到 8 倍光罩极限。这一面积规模恰好能够满足当前 Rubin Ultra 的封装设计需求，为其性能提升奠定了基础。

图表7：台积电 CoWoS 中介层 Roadmap



来源：anandtech、国金证券研究所

图表8：MI300X 使用四个 I/O Die



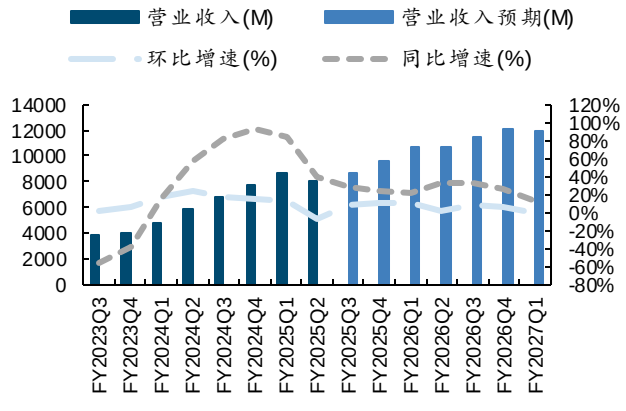
来源：Chips and Cheese、国金证券研究所

然而，从目前掌握的信息来看，Rubin Ultra 并未采用一整块达到 8 倍光罩极限的中介层设计，而是借鉴了类似于 MI300 I/O die 的设计思路。具体而言，MI300X 并非仅采用单个 I/O die，而是采用了四个 I/O die，并通过超高带宽的互连技术连接在一起。其计算 die 堆叠在 I/O die 之上，而 I/O die 又堆叠在中介层之上，形成了一个三层堆叠结构。我们认为，Rubin Ultra 选择这一相对保守的封装设计，某种程度上也反映出当前大面积 CoWoS 封装的挑战依然不容小觑。至于 Rubin Ultra 是否会面临与 Blackwell 类似的设计问题，目前尚存在不确定性，仍需进一步观察。

美光 FY25Q2 财报：消费级闪存重回增长，数据中心 SSD 采购放缓

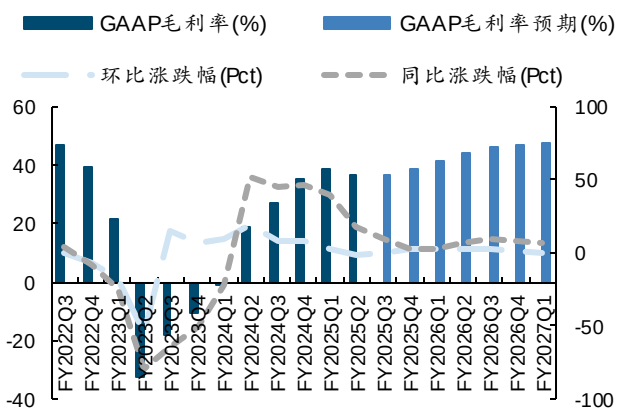
美光科技于 3 月 20 日盘后发布了其 FY25 Q2 财报。报告显示，公司本季度实现营业收入 80.53 亿美元，同比增长 38.27%，环比下降 7.53%。其中，DRAM 业务收入环比下滑 4%，公司解释称这是由于 DRAM 产品价格上涨的影响被 bit 出货量的减少所抵消。NAND 业务收入环比下滑 17%，主要受到价格下降的拖累。本季度公司 GAAP 毛利率小幅下滑 1.64 个百分点至 36.79%，主要由于 NAND 业务中低端消费级产品出货占比的提升所致。这一表述与公司对 NAND 业务收入环比下滑的解释相互呼应。

图表9：美光营业收入



来源：Reuters、国金证券研究所

图表10：美光 GAAP 毛利率



来源：Reuters、国金证券研究所

上周我们曾指出，美光本季度财报是观察数据中心 SSD 需求景气度的重要指标。鉴于美光在该领域具备深厚的技术积累，本季度公司存储业务部门（SBU）收入环比下滑 20%，主要由于数据中心客户在经历数个季度的高强度采购后，本季度需求出现回落。结合过去一个月上游 TLC 和 QLC Flash Wafer 报价的走势来看，我们推测下游数据中心客户对价格上涨的承接能力有限，短期内仍处于观望状态。与此同时，公司在本季度也表示，将继续保持 NAND 领域相对较低的稼动率，以控制供应节奏。因此，尽管短期内 NAND 价格呈现较为显著的上涨，但市场景气度能否持续回暖仍有待进一步观察。

图表11：本周 TLC 大容量 NAND Flash Wafer 价格涨幅显著，QLC 涨幅环比下滑 4Pct

Flash Wafer周度涨跌幅

料号	2025年07周	2025年08周	2025年09周	2025年10周	2025年11周	2025年12周
QLC 1Tb	0.00%	0.00%	0.00%	0.00%	6.67%	2.08%
TLC 1Tb	0.00%	-1.85%	-1.89%	0.00%	1.92%	3.77%
TLC 256Gb	0.00%	10.34%	0.00%	0.00%	0.00%	18.75%
TLC 512Gb	0.00%	0.00%	-3.39%	0.00%	3.51%	0.00%

Flash Wafer月度涨跌幅

料号	2024年10月	2024年11月	2024年12月	2025年1月	2025年2月	2025年3月
MLC 128Gb	0.00%	0.00%	0.00%	-2.21%	0.00%	0.00%
MLC 256Gb	0.00%	0.00%	0.00%	-11.58%	0.00%	0.00%
MLC 32Gb	0.00%	0.00%	0.00%	8.80%	0.00%	0.00%
MLC 64Gb	0.00%	0.00%	0.00%	2.17%	0.00%	0.00%
QLC 1Tb	-12.07%	-1.96%	-6.00%	-4.26%	0.00%	8.89%
SLC 16Gb	0.00%	0.00%	0.00%	-0.64%	0.00%	0.00%
SLC 1Gb	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SLC 2Gb	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SLC 4Gb	0.00%	0.00%	0.00%	3.63%	0.00%	0.00%
SLC 8Gb	0.00%	0.00%	0.00%	2.78%	0.00%	0.00%
TLC 128Gb	0.00%	0.00%	0.00%	-14.29%	0.00%	0.00%
TLC 1Tb	-10.45%	-1.67%	-5.08%	-3.57%	-3.70%	5.77%
TLC 256Gb	-3.78%	0.00%	3.57%	-9.31%	21.67%	18.75%
TLC 512Gb	1.43%	0.00%	-1.56%	-15.08%	6.54%	3.51%

来源：中国闪存市场、dramexchange、国金证券研究所

Agentic AI 推动加速卡容量进一步提升，海力士正式发布 12 层 HBM4

GTC2025 上英伟达提出 Agentic AI 将是一个能够使多个模型协同工作的人工智能模式，叠加思考时的计算量，我们认为将对加速卡存储容量和带宽提出更高的要求。相应的，本次会议上发布的 Rubin Ultra 所搭载的存储容量大幅提升，单卡达到 1024GB，带宽同时也提升到 32TB/s。

图表12：加速卡 HBM 容量和带宽持续提升

GPU	公司	发布时间	发货时间	制程	HBM类型	HBM容量(GB)	HBM带宽
MI325X	AMD	2024.10.10	2024Q4	> AIDs: TSMC 6nm FinFET > Die: TSMC 5nm FinFET	HBM3e	256	6TB/s
H100 SXM	NVIDIA	2022.03.22	2022Q3	TSMC 4N Customized for NVIDIA	HBM3	80	3352GB/s
H200 SXM	NVIDIA	2023.11.13	2024Q2	TSMC 4N Customized for NVIDIA	HBM3e	141	4.8TB/s
B200	NVIDIA	2024.03.18	2024Q2	TSMC 4NP Customized for NVIDIA	HBM3e	192	8TB/s
B300	NVIDIA	2024.03.18	2024Q3	TSMC 4NP Customized for NVIDIA	HBM3e	288	8TB/s
Rubin	NVIDIA	2025.03.18	2026H2	N3P	HBM4	288	13TB/s
Rubin Ultra	NVIDIA	2025.03.18	2027H2	N3P	HBM4e	1024	32TB/s

来源：英伟达、AMD、Semianalysis、Trendforce、anandtech、国金证券研究所

SK hynix 近日也发布了其最新一代 HBM 产品 HBM4，从技术角度来看，SK hynix 的 12 层 HBM4 具备业界最高容量与带宽，首次突破 2TB/s 带宽大关，比 HBM3E 提升 60%，在 AI 大模型、HPC（高性能计算）等数据密集型场景中将显著提升性能表现。同时，HBM4 的 36GB 单体容量创新纪录，结合其先进的 MR-MUF 工艺，有效避免芯片翘曲并提升散热稳定性，为大规模部署奠定了可靠基础。我们认为随着 Agentic AI 到来，HBM 该细分赛道仍将维持较高的景气度。

英伟达 AGX Hyperion 自动驾驶平台

GTC 2025 上，英伟达发布 Nvidia Halos 以保障自动驾驶安全，同时通过 Omniverse 与 Cosmos 构建世界模型与模拟驾驶完成自动驾驶算法的验证。大会上，英伟达宣布通用汽车将利用英伟达的计算平台与硬件，包括 Omniverse、Cosmos 与 Nvidia AGX 硬件，构建从自动驾驶到智能座舱的 AI 体系。

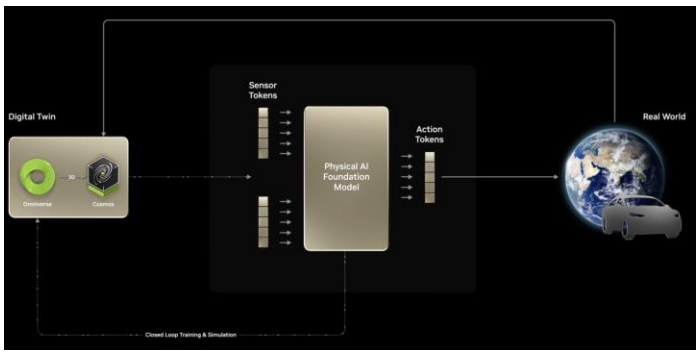
图表13：Omniverse+Cosmos 训练下的自动驾驶

来源：英伟达、国金证券研究所

NVIDIA DRIVE AGX Hyperion 是一款完整的自动驾驶平台，集成了整套传感器架构、高性能 AI 计算能力和强大的软件栈，可加速自动驾驶汽车的开发和部署。作为 NVIDIA 针对自动驾驶汽车开发的三个计算平台解决方案的一部分，DRIVE AGX 可与用于 AI 模型训练的 NVIDIA DGX 和用于仿真和验证的 NVIDIA Omniverse(搭配了 Cosmos) 协同工作。NVIDIA Halos 作为端到端安全的基础，可将硬件、软件、工具与模型相结合，以保护从云端到汽车的整个自动驾驶汽车堆栈。

Omniverse 与 Cosmos 的合作帮助自动驾驶训练速度加快。Cosmos 作为世界模型在其通过 Omniverse 生成各种各样的驾驶场景数据训练后在成为自动驾驶端到端模型。精调过的模型在自动驾驶中的决策也可以在 Omniverse 生成的各种场景中得到验证与反馈。Cosmos 可以生成同一场景的不同变化，包括天气、光照和地理环境等因素，进一步改善恶劣环境下数据的感知与自动驾驶的决策。

图表14：英伟达自动驾驶方案



来源：英伟达、国金证券研究所

图表15：Omniverse 生成各种类型的驾驶场景

来源：英伟达、国金证券研究所

NVIDIA Halos 是一个先进的自动驾驶汽车安全系统，由硬件/软件组件、工具、模型和设计原则构成，可将它们配合使用，保护从云端到汽车的端到端自动驾驶汽车堆栈。Halos 从英伟达自身的自动驾驶汽车堆栈开发演化而来，可对现有行业安全实践进行补充和增强，包括遵循法规和标准化框架。

虽然在 GTC Keynote 2025 上没有提及，但在今年 1 月 CES 上，英伟达宣布新一代基于 Blackwell 架构打造的高性能 DRIVE AGX Thor 芯片将在今年上半年推出。Thor 预计将提供高达 2000 TOPS 的算力，相比目前 Orin 芯片的 254 TOPS 算力了提升了约 8 倍。

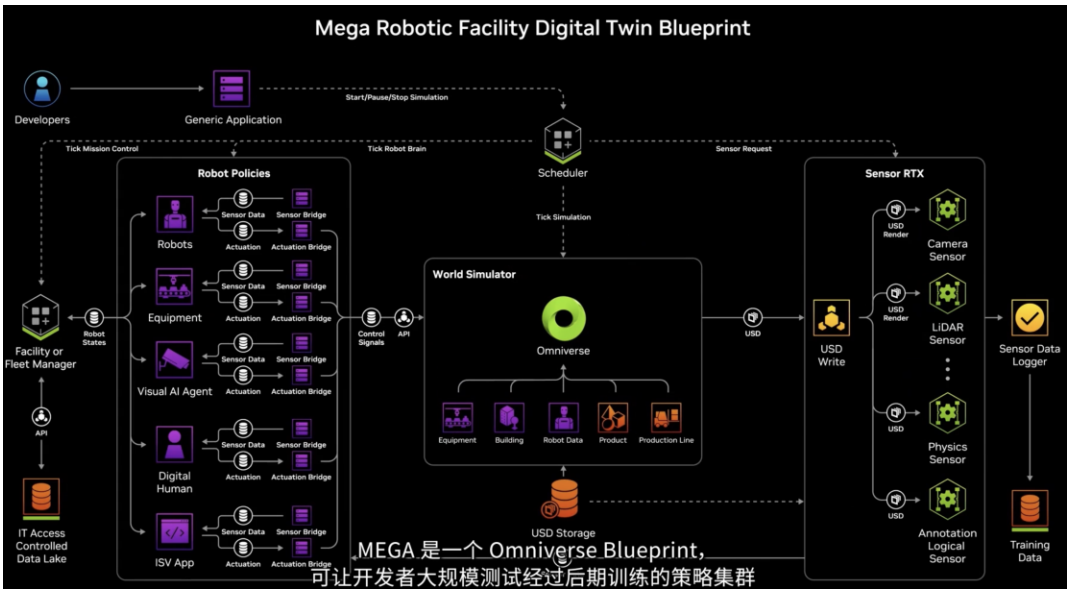
当前自动驾驶痛点包括训练数据不足、决策规划的安全性不足、传感器精度不足、系统安

全性不足、车载芯片算力不足、高精地图缺失等等。英伟达 Omniverse+Cosmos 方案下可以帮助解决训练数据、决策规划安全性的不足，Halos 将保护 Hyperion 系统级的安全，新一代 Thor 芯片也将解决短期内芯片算力不足的情况。我们认为自动驾驶发展将会进一步加速，与英伟达有合作的企业将会率先受益。除了合成数据帮助训练自动驾驶算法之外，真实世界驾驶数据也将帮助算法训练。国内安防企业通过摄像头已采集大量车辆驾驶数据，我们认为这些公司也将受益。

机器人大脑也将受益于 Omniverse+Cosmos

GTC 2025 上除了自动驾驶之外，黄仁勋也着重介绍了 Omniverse+Cosmos 生态另一个重要应用——机器人。与自动驾驶类似，机器人也需要世界模型与大量真实世界数据来训练。Omniverse+Cosmos 的世界模型+合成数据训练可以加速机器人大脑智能发展。再加上配套英伟达 Issac 机器人库进行后期策略训练，机器人可以通过强化学习进一步提升智能度。Mega 可以让开发人员同时研究复数机器人策略，使机器人可以作为系统工作。

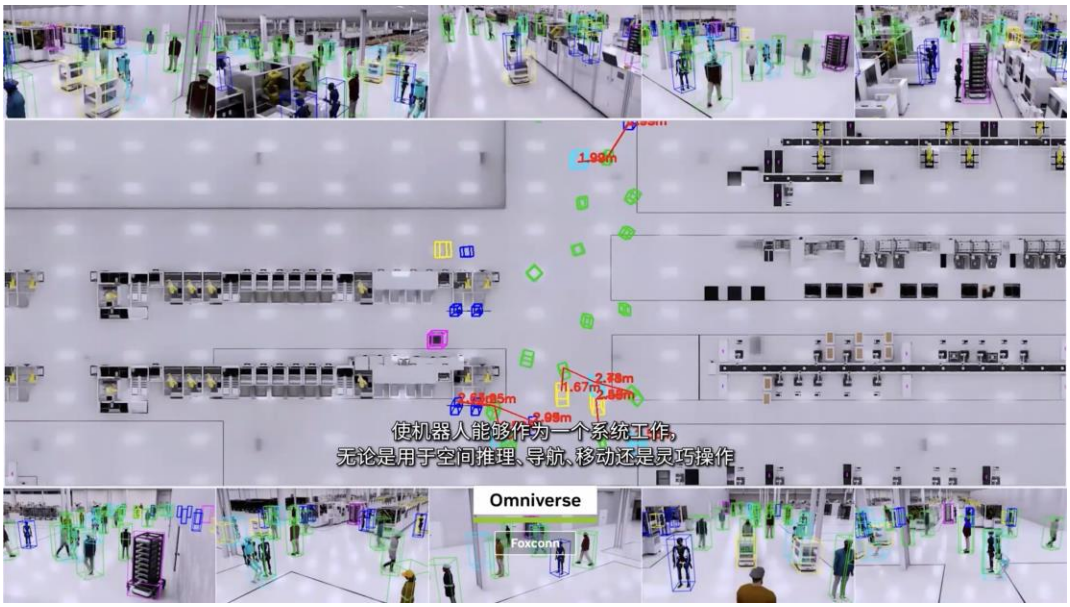
图表16：Mega 帮助复数机器人协同训练



来源：英伟达、国金证券研究所

富士康正在 Omniverse 孪生出的 Blackwelk 生产设施中测试异构机器人

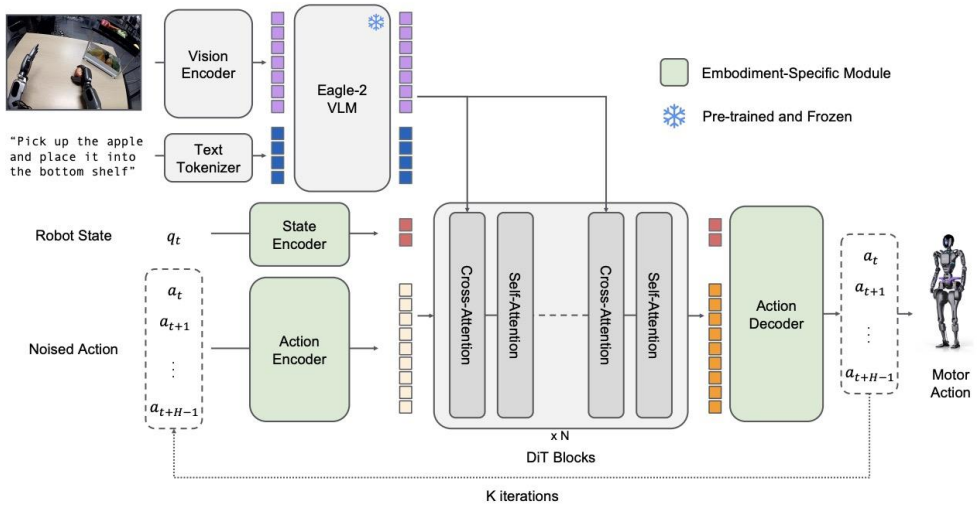
图表17：富士康异构机器人测试



来源：英伟达、国金证券研究所

英伟达也公布了开源的 Groot N1 模型。GROOT N1 是一种视觉-语言-动作（Vision-Language-Action，VLA）模型，具有双系统架构。视觉语言模块（系统 2）通过视觉和语言指令理解环境；随后，扩散变换模块（系统 1）实时生成流畅的动作指令。这两个模块紧密耦合，并联合端到端训练。该模型包含一个视觉-语言主干，用于对语言和图像输入进行编码，以及一个基于扩散变换器的流匹配策略，用于输出高频动作指令。英伟达采用 NVIDIA Eagle-2 VLM 作为视觉语言主干模型。GROOT-N1-2B 模型总共包含 22 亿个参数，其中视觉语言模型（VLM）占 13.4 亿。

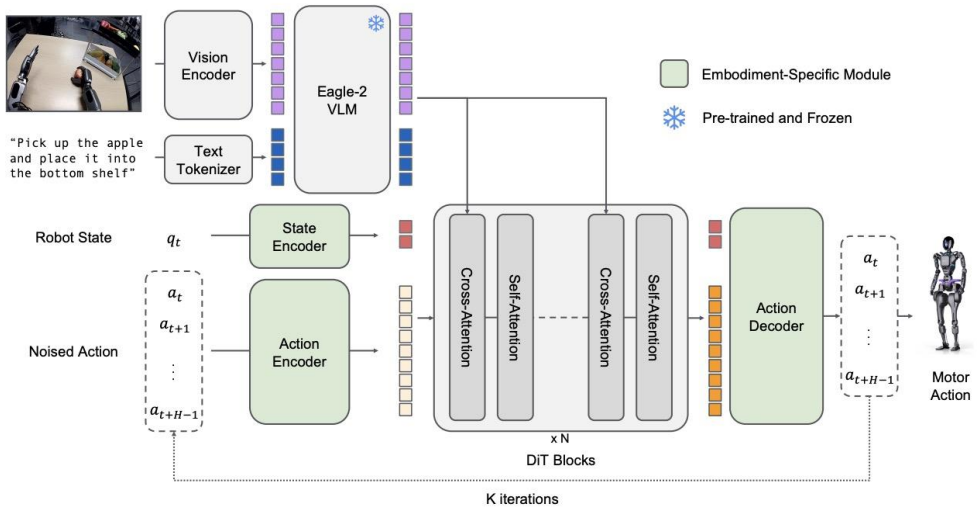
图表18：Groot N1 架构



来源：英伟达、国金证券研究所

RoboCasa 中任务“转动水龙头”为例，在 100 条数据场景下，Diffusion Policy 成功率仅为 11.8%，而 GROOT N1 为 42.2%。Diffusion Policy 经常无法理解任务语义，而 GROOT N1 在低数据环境下具有显著优势，体现了预训练的效果。

图表19：Groot N1 架构



来源：英伟达、国金证券研究所

目前在机器人制造方面，国内产业链更为成熟，企业产能、供应链管理相比美国企业更具优势，但在机器人大脑、小脑智能上相比有英伟达芯片支持的美国企业仍有差距。我们认为英伟达 Omniverse+Cosmos+Groot N1 生态叠加 Thor 平台今年上线将会加速机器人具身智能发展，看好产业链受益。

风险提示

1. 芯片制程发展与良率不及预期：半导体工艺的发展面临诸多挑战，主要包括技术瓶颈、良率提升难度、研发成本高企以及供应链不确定性等问题。随着工艺节点微缩变得愈发复杂，先进制程的实现难度和成本不断攀升，可能导致量产延迟，甚至影响产品性能和成本控制。此外，地缘政治风险和出口管制可能扰乱供应链，进一步拖累产能扩张。

2. 中美科技领域政策恶化：中美在 AI 领域竞争激烈，美国限制先进芯片和半导体对中国的出口，随着竞争的加剧，未来可能会推出更严格的限制政策，限制国内 AI 模型的发展。

3. 智能手机销量不及预期：智能手机销量与产品本身质量关系紧密，若产品本身有缺陷则智能手机销量可能收到影响。同时宏观经济变化也有可能导致消费者消费意愿发生变化从而影响智能手机销量。

特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

上海
电话：021-80234211
邮箱：researchsh@gjzq.com.cn
邮编：201204
地址：上海浦东新区芳甸路 1088 号紫竹国际大厦 5 楼

北京
电话：010-85950438
邮箱：researchbj@gjzq.com.cn
邮编：100005
地址：北京市东城区建内大街 26 号新闻大厦 8 层南侧

深圳
电话：0755-86695353
邮箱：researchsz@gjzq.com.cn
邮编：518000
地址：深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806

【小程序】
国金证券研究服务

【公众号】
国金证券研究